DOCUMENT RESUME

TE 000 502

ED 022 748

By-Purves, Alan C.

DESIGNING THE BOARD'S NEW LITERATURE ACHIEVEMENT TEST.

Pub Date 68

Note-6p.; Adapted from an address given at the Annual Convention of the National Council of Teachers of English (Honolulu, Hawaii, Nov. 1967).

Available from-College Entrance Examination Board, Publications Order Office, Box 592, Princeton, New Jersey 08540 (Subscription rate, $2.00 per year. Single copy, $0.50).

Journal Cit-College Board Review; n67 p16-20 Spring 1968

EDRS Price MF-$0.25 HC Not Available from EDRS.

Descriptors-CRITICAL READING, CRITICAL THINKING, *EDUCATIONAL TESTING, ENGLISH INSTRUCTION, LITERARY ANALYSIS, LITERARY CRITICISM, LITERARY DISCRIMINATION, LITERARY GENRES, *LITERATURE, LITERATURE APPRECIATION, MULTIPLE CHOICE TESTS, *OBJECTIVE TESTS, *TEST CONSTRUCTION, *TESTING PROBLEMS

Identifiers-*College Board Achievement Tests

This article describes the problems that the College Entrance Examination Board's Committee of Review for the Examinations in English encountered in creating a fair, objective, hour-long literature achievement test which would meet four objectives--to measure the breadth of a student's reading, his understanding of that reading, his response to literature, and his critical skills. Sample multiple-choice questions on Louise Bogan's poem "Medusa," illustrate (1) the kinds of questions chosen by the Committee, (2) the eight skills that the questions are designed to test, (3) the reasons for choosing these types of questions, and (4) the pattern of questioning which moves from testing sharpness of observation to depth of understanding. (JS)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

College Board Review

SPRING 1968, NO. 67

News of the College Board, 3

Letters, 3

How can we make testing more relevant?, 5
by William W. Turnbull

Upward Bound at Bowdoin: Who's teaching whom?, 11
by Daniel Levine and Herbert R. Coursen Jr.

Designing the Board's new literature Achievement Test, 16
by Alan C. Purves

Do most students need 'time out' between school and college?, 21
by Howell F. Nomer

New role for the student aid officer: 'resourceful uncle', 24
by John W. Harris

Quotes, 28

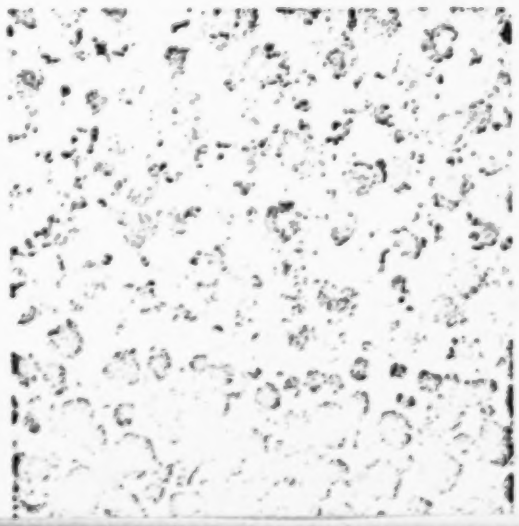
College Board Review is published quarterly
(fall, winter, spring, and summer) by the
College Entrance Examination Board,
475 Riverside Drive, New York, New York 10027
Address editorial contributions and inquiries to Robert L. Brenner, Senior Editor.

Subscription rates: $2 per year; 3 years, $5; 5 years, $7; single copy, $.50.

Address subscription orders
and orders for other College Board publications to
College Entrance Examination Board,
Publications Order Office, Box 592, Princeton, New Jersey 08540.

Copyright © 1968 by College Entrance Examination Board.
All rights reserved. Library of Congress Catalog Card Number: 56-46771.
Printed in the United States of America.

ED022748
TE 000 502



Illustrations: Our Spring cover is by Ian Reid. Vincent Malta created the frontispiece and illustrated the articles beginning on pages 5 and 11. Stanley Wyatt illustrated those beginning on pages 16, 21, and 24. The cartoon on page 14 is Henry Martin's.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY College Entrance Examination Board TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U. S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

BY ALAN C. PURVES

Designing the Board's new literature Achievement Test

To construct a nonessay literature test that measures both sharpness of observation and depth of understanding, yet is still fair to students of varied backgrounds, is less easy than you may have guessed.

This May, the College Board offered for the first time a one-hour, multiple-choice Achievement Test in English literature. In effect, the new test is an outgrowth of the recommendations of the Committee of Review for the Examinations in English, appointed by Board President Pearson in 1964 to review the Board tests in English in the light of the findings of the Commission on English and the testing needs for admissions and placement. A major recommendation of this committee was that the Board's English examinations should reflect the tripartite nature—language, literature, and composition—of most English curriculums today.

Work on the test began in October 1965, when Board members authorized the Board's Committee on Examinations to develop an experimental two-hour test of both English composition and literature. The Committee reported in March 1967 that preliminary experiments had shown the combined two-hour test to be neither psychometrically nor administratively feasible, but that the objective portions of the literature section seemed usable as a one-hour test. With the approval of the Board's trustees, the first administrations of the new test were scheduled for this May and for twice next year.

In this article, one of the test's designers details the "gripping saga" of how the literature committee of the Committee on Examinations created the new test: what objectives were considered, how they were evaluated, what kinds of questions were finally selected, and why.*

Alan C. Purves, a former member of the test development division of Educational Testing Service, is now an associate professor of English at the University of Illinois. His article is slightly adapted from his address before the members of Pi Lambda Theta at the annual convention of the National Council of Teachers of English, held last November in Honolulu.

The literature committee was originally charged by the College Board's Committee of Review for the Examinations in English with creating a test to meet four objectives: 1) to measure the breadth of a student's reading; 2) to measure his understanding in some depth of works he has studied and read; 3) to measure his response to literature, jointly affective and evaluative; and 4) to measure his ability to use whatever critical skill he has on texts unfamiliar to him.

These four objectives include most of those listed in statements on secondary school literature teaching. Omitted are the attitudinal and the broad, humanistic objectives which are long-term goals rather than immediate rewards. Each of the four presents certain measurement problems, particularly within the hour-long format of College Board Achievement Tests. My hope is that this gripping saga of the committee's struggle with these problems will be of use to the creator of classroom tests.

The first objective, to measure the breadth of a student's reading, seems easiest, particularly with multiple-choice questions. What could be simpler than: Duncan is Macbeth's (A) son; (B) brother; (C) king; (D) murderer; (E) conqueror? Such a question can separate those who remember the play from those who do not. It is unimpeachable. And when taken by itself, it is unfair.

Macbeth is the most-taught work in American high schools—about 70 percent of them include it in their curriculums. But what about the other 30 percent? When you go beyond *Silas Marner, Romeo and Juliet, Huckleberry Finn,* and similar pedagogical bestsellers, you will find that not more than about 10 percent of the schools assign any single title. Particular short stories and lyric poems are used by even fewer schools. To give every student a fair shake, a test of literary breadth should include at least 100 questions, and even then the examiner gains less an index of achievement than of mere acquaintance.

The second objective, "to measure understanding in depth" of works studied by the student, poses a different sort of problem. Such a measure requires an essay answer, because a set of multiple-choice questions cannot be made up for every book a student might have read, and a universal series of such questions is absurd. Essay questions do exist in the Advanced

*Committee members were Laurence A. Michel Jr., State University of New York at Buffalo, chairman; Arthur A. MacConochie, Norfolk Academy, Norfolk, Va.; Martin L. Parry, John Burroughs School, St. Louis, Mo.; R. Christine Royer, Barnard College; William T. Stafford, Purdue University; Nell H. Thomas, Greenville High School, Miss.; and Graham C. Wilson, San Jose State College.

Placement English Examination, but even they run into practical snags that would take on immense proportions in a really large-scale examination.

The first snag is that of differing interpretations. If, in all honesty, some benighted candidate wrote an interpretation of *Macbeth* which made Duncan a symbol of France, Macbeth of Ho Chi Minh, and Macduff of Henry Cabot Lodge, most responsible graders would color him dense. But it might be the interpretation he was taught. The second snag is, once again, that of variety. Few Advanced Placement students choose Leon Uris as their author of literary merit; he isn't assigned in Advanced Placement classes. But his work is often assigned in other classes, and a student cannot be blamed for writing about the theme of man and society as exemplified, for instance, in *Exodus*.

The third snag, that of sampling, is the biggest one. The Advanced Placement Examination is three hours long. This means the candidate can write at least two essays of this sort as well as answer other questions. An inappropriate choice, a misunderstanding of the question, or a dubious interpretation in one essay can be atoned for in the other. In an hour-long essay test, however, the number of targets at which to shoot is limited, and the candidate cannot demonstrate his ability as fully as he might.

Mastery and response: untestable?

In addition to these three snags, there is a theoretical criticism of the importance of the goal itself. It maintains that you teach a particular work —*Macbeth*, for instance—less so that students may become masters of *Macbeth* than that later they may be equipped to deal with *Hamlet*. The goal of understanding in depth, then, is only a way station to a larger goal. In any case, the committee decided that the circumstances under which the new test was to be introduced militated against measuring the objective of understanding in depth alone. However, a brief essay on a book of the candidate's choice still could be paired with objective three or four.

Objective three, "to measure his response to literature, jointly affective and evaluative," deals with one of the most important goals of instruction in

literature. If students aren't moved by books and have not developed a taste for them, then English teachers have failed. Everything else is merely a means to this goal, the achievement of which will determine the reading habits and the very being of the student. How then to measure response? Should the candidate report on his enjoyment of what he is reading? Should his taste be measured against some norm? Should we take an electroencephalograph?

The answer to the question of measuring affective response is that we don't know yet how to do it well enough. In part, we do not trust the students to be honest; often enough we ourselves aren't honest, particularly when we teach something for the twentieth time—*Silas Marner*, for instance —simply because the school board won't buy another set of books until the old ones wear out. Besides, affective responses differ in nature, and no single criterion exists. We want students to be caught up in books, but at the same time we want them to be detached observers and interpreters. The balancing point is hard to find.

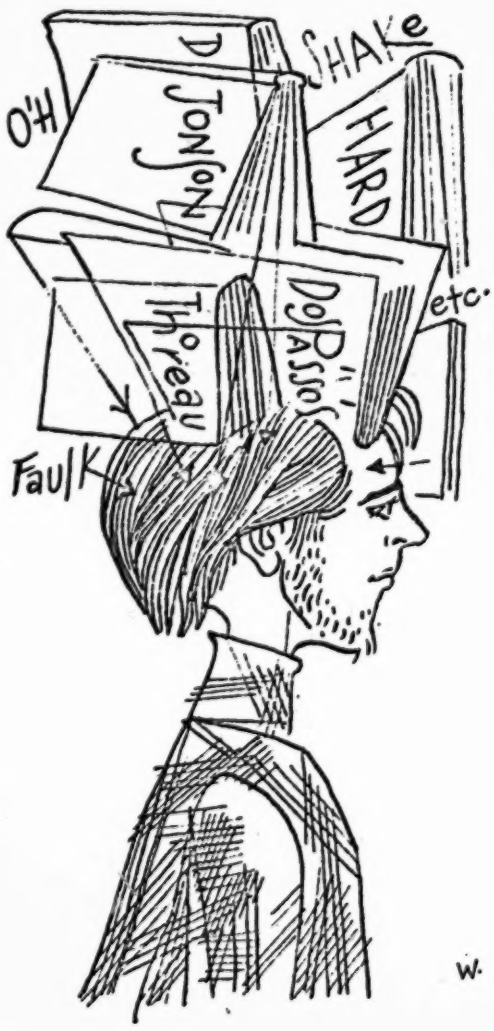
How about testing for a measure of taste? One might use 20 passages—10 judged good, 10 bad, by a panel of teachers—and ask students to rank them. Experiments with this sort of test have shown, though, that more often than not the students make poor choices. Besides, what teacher would be willing to do more than make a rough distinction between Shakespeare and Neil Simon, between William Faulkner and Mickey Spillane? The middle ground on which John O'Hara and John Updike reside is full of quicksands. A taste test is intriguing dynamite but, unfortunately, inappropriate as a College Board Achievement Test. A measure of the *premises* of taste might have a better chance of success, but there are still problems. Whether a poem is good primarily because of its form, its theme, or its effect on the reader is moot: what counts is the persuasiveness of the argument for using any one of these criteria in a particular case. In effect, such a test would be more a test of composition than of literary skill.

The last objective, "to measure his ability to use whatever critical skill he has in texts unfamiliar to him," also presents problems. Critical skill may be said to consist of the ability to perceive the parts and the relationship of parts to each other and to the whole of a work; the ability to devise a coherent interpretation of the work; the ability to perceive the effect of the work on the reader's response; and the ability to make a reasoned judgment about the work. The problem in measuring this goal is the tentativeness you must assume in asserting anything definitive about a person.

With some degree of assurance, you can say that a person who can answer six subtraction problems using three columns can answer most subtraction problems. But assurance melts away when you discover that a student can answer six questions on imagery, rhythm, substance, tone, structure, and point of view in a lyric by William Butler Yeats. Can he do the same with

another poem by Yeats, let alone with one by a poet like Robert Herrick or Ben Jonson? We know that students who have cut their critical teeth on John Donne, Andrew Marvell, and the early T. S. Eliot are often incapable of talking or writing about Walter Landor, William Cowper, or Alfred, Lord Tennyson. This is not to say that they are not good students of literature, but that they have become so absorbed in the explication of difficult works that they have trouble appreciating simplicity. Other students may be put off by certain words in a poem. After all, that is what literature is—a quicksilver art that changes with each reading without losing its weight and mass.

Anyone measuring this objective, therefore, must consider the uniqueness of the work and the response to it, and the tentativeness of any assertion about achievement. Yet this objective, the developed ability to apprehend a new literary work, is the chief one underlying most literature curriculums—be they historical, thematic, or generic. Further, the problems of measurement are ones that can be borne more easily than can the problems involved in each of the other objectives. For this reason, as well as for the reason that a measure of critical ability would influence secondary school curriculums positively and would not hamstring them, as might tests of literary acquaintance or tests on set books, the literature committee decided that the Achievement Test in literature should "seek knowledge about a student's ability to comprehend, analyze, and evaluate literary works. It will focus on passages in the various genres, drawn primarily from English and American literature."



The eight skills finally chosen

It was decided that the test questions should seek to measure a student's ability: 1) to paraphrase parts of the work or summarize the whole work; 2) to comprehend the structure of the text; 3) to comprehend language and style; 4) to comprehend rhetorical and literary devices; 5) to comprehend the ways by which structure and language, and rhetorical and literary devices enhance and even create the meaning and form of a work; 6) to classify a text by genre, tradition, or period; 7) to understand allusions to common figures or symbols in mythology, literature, and folklore; and 8) to deal with such general aspects of literary study as theme, history, and the writer's art.

The last of these abilities was included so that the committee might feel free to add an essay portion to the test should it deem such an addition worthwhile. For the foreseeable future, however, the test will consist of multiple-choice questions, all of them directed at specific passages—prose, poetry, or drama—presumably unfamiliar to the candidate. In the hour-long test, there will be from 6 to 12 passages varying in length, genre, form, and historical period. This number of passages should provide a sufficient aggregate to enable one to assert that a candidate who does well on the test can handle not any text that might come along, but a goodly number of them. Such a candidate certainly should be acknowledged as a competent reader of literature, and therefore that much more eligible for admission to a liberal arts college.

What of the questions, however? Can any *searching* questions be cast into the multiple-choice mold? The answer depends on what you mean by "searching." Certain limitations are imposed on the questions. Let me take as an example the sample questions on "Medusa" by Louise Bogan, in the box on the next page (asterisks indicate correct answers).

These 10 questions represent the sort that will be asked on the new test. The first asks the source of an allusion, the second the form of the work. Both require factual information that the student should be able to bring to the poem. (Other appropriate factual questions might be about approximate date, or in some cases the probable author.) Questions 3, 4, and 5 are also relatively factual, but deal with internal aspects of the poem. Numbers 3 and 4 ask about the effect of the language and ask the student to reject an implausible connotation or an unimportant point about structure. In question 5, the student is asked to contrast the details in two parts of the poem. The last of these can be verified empirically: the first two depend on the student's knowledge of semantics and of the concept of organicism in a work of art (that certain aspects of linguistic form are coincidental and others are meaningful).

The first four questions, then, are relatively straightforward and empirically verifiable, given a basic consensus about the nature of poetic language and form. But you may argue that these are not searching, and you would be right; yet to answer them correctly is to display a basic knowledge of what literary language is all about.

Three measures of discernment

Questions 6, 7, and 8 move beyond this point. They ask students to make summary statements about the relationships between parts of the poem (question 6), about the tenor of a metaphor (question 7), and about the functional effect of stylistic devices (question 8). The first of these questions calls upon an ability to make generalizations about the content, the second to comprehend the relationship of resonant language to the work as a whole, and the third to relate a stylistic device—in this case aberrant versification—to the pace of the poem. This last question goes a bit further than the others in that it asks students to reject various overinterpretations of stylistic devices. It is perhaps a dangerous question, but I think its incorrect options can all be dismissed as peripheral at best, misguided at worst.

These three questions, then, go beyond the factual and ask students to read discerningly, keeping one eye on the poem as a whole and the other on certain of its facets. They do not go beyond that which is verifiable by reference to the verbal context. (Other questions might have been asked: on the relation between Medusa and a

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY Farrar, Straus, and Giroux, Inc. TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U. S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

MEDUSA

I had come to the house, in a cave of trees,
Facing a sheer sky.
Everything moved,—a bell hung ready to strike,
Sun and reflection wheeled by.

When the bare eyes were before me
And the hissing hair,
Held up at a window, seen through a door.
The stiff bald eyes, the serpents on the forehead
Formed in the air.

This is a dead scene forever now.
Nothing will ever stir.
The end will never brighten it more than this,
Nor the rain blur.

The water will always fall, and will not fall,
And the tipped bell make no sound.
The grass will always be growing for hay
Deep on the ground.

And I shall stand here like a shadow
Under the great balanced day,
My eyes on the yellow dust that was lifting in the wind,
And does not drift away.

Reprinted from *Collected Poems* by Louise Bogan by permission of Farrar, Straus & Giroux, Inc. Copyright 1954 By Louise Bogan.

1. Medusa is the name of
(A) a sea nymph
(B) the muse who was the inspiration of lyric poets
(C) the girl who led Theseus through the labyrinth
(D) a woman who turned those who looked at her to stone
(E) the girl who married Jason after helping him find the Golden Fleece

2. This poem is written in
(A) free verse
(B) heroic couplets
(C) sonnet form
(D) irregular quatrains
(E) blank verse

3. In stanza 2, depicting Medusa's eyes as "bare," "stiff," and "bald" does all of the following *except*
(A) indicate that the speaker only imagines her effect on him
(B) prepare for the description of the speaker's state in stanzas 3-5
(C) emphasize her inhuman appearance and its effect on the speaker
(D) create a contrast with "moved" and "wheeled" in stanza 1
(E) show that her eyes are a prime force behind her spell

4. The contrast in effect between stanza 2 and stanzas 3, 4, and 5 is produced by all of the following devices in stanza 2 *except* the
(A) single-syllable rhyme, "air" in line 6 and "hair" in line 9
(B) lack of a finite verb in the sentence in lines 8-9
(C) lack of a main clause in the sentence in lines 5-7
(D) four short phrases in lines 7-8
(E) word "When" line 5

5. Which of the following is true of the description in lines 16-17 but not of the description in lines 10-15?
(A) Lines 16-17 contain references to natural events.
(B) There is a lack of action in lines 16-17.
(C) Events described in stanza 1 are repeated in lines 16-17.
(D) In lines 16-17, there is a possibility that Medusa's spell is incomplete.
(E) In lines 16-17, there is a reference to the future like the references in stanza 2.

6. The difference between the scene in stanza 1 and that in stanzas 3, 4, and 5 is primarily the difference between
(A) joyful events and sad events
(B) a realistic scene and a scene in a play
(C) a scene on earth and a scene in hell
(D) action that did happen and action that will happen
(E) action and arrested action

7. By comparing himself to a "shadow" (line 18), the speaker indicates that
(A) in this scene his own reality is dubious
(B) his consciousness has been destroyed by his experience
(C) the sky has grown brighter than it was when the poem began
(D) he thinks he will disappear when the daylight finally ends
(E) he has died and become a ghost because of Medusa's spell

8. What is the function of the extra line and more rapid movement in stanza 2?
(A) To suggest that Medusa was a vigorous and active creature
(B) To indicate that the speaker moved quickly when he was frightened
(C) To focus attention on the central experience of the poem
(D) To show that the central experience lasted longer than others in the poem
(E) To help the reader to see that the poem is like an old ballad

9. The events in the first stanza are told in the past tense; the events in the third, fourth, and fifth stanzas are told in the present and future tenses. Which of the following is the best inference one can draw from this fact?
(A) The first scene happened in the speaker's past; the second scene is an attempt to relate what happened after the speaker left.
(B) The first scene happened when the speaker was alive; the second scene occurs after death.
(C) The first scene is perceived by the speaker; the second scene is perceived by Medusa.
(D) The first scene is real; the second scene is a metaphoric interpretation of it.
(E) The first scene occurred in the speaker's life; the second scene occurs in his mind.

10. The poem as a whole deals with
(A) encountering and killing the mythical Medusa
(B) experiencing an unnamed shock as powerful as Medusa's spell
(C) meeting someone as horrible to look at as Medusa was
(D) enjoying the idea that the myth of Medusa is not true
(E) thinking about how Medusa would look if she were real

On a multiple-choice test, interpretive questions must be limited in their scope and in their penetration or risk being untenable.

house; on the function of "stir," "brighten," and "blue"; on the summary nature of the word "balanced" in line 19; or on the mood of the speaker in the last two stanzas.) All of these call for accurate description of the words in the context of the poem.

How do you test interpretation?

It might be argued that objective questions cannot go beyond this point, but questions 9 and 10 do, since both ask for *interpretations* of the poem or its parts. Question 9 asks for the best inference to be drawn from an observation of the poem's structure. (Option B can't be totally rejected, perhaps, but E is better, if only because poetic speakers do not usually speak from the grave. Options A and C are clearly unsubstantiated, and option D violates much of the logic of the poem.) In a similar way, question 10 asks for the best summary of the poem. (Clearly option B is most consonant with the statement and mood of the poem; only option C has a plausible ring, but it is conjectural at best and finally insupportable because there is really little enjoyment in the poem.)

Of these two questions, 9 is the more searching, but it would seem that a question like 10 could be made more penetrating and still have one clearly acceptable option. For instance, suppose that the word in option D was "contemplating" instead of "enjoying" the idea that the myth of Medusa is not true. It might not be as good an answer as B, but the word contemplating is ambiguous enough to render the option acceptable. The same might be said of option D if it were changed to "encountering and being petrified by the mythical Medusa." This latter statement, while not the poem's subject matter, does underlie it. In any case, the wrong options for this sort of question must be clearly inadequate to the poem.

Thus, questions of interpretation must be limited in scope and penetration. The following two questions provide a good example of these limits: Both are good discussion questions but would be untenable on a multiple-choice test.

The first one asks: Which of the following does *not* describe the overall progression of the poem? (A) from ignorance to knowledge; (B) from life to death; (C) from activity to paralysis; (D) from time to timelessness; (E) from happiness to despair? This is a favorite sort of question in literature, asking the student to see movement in the poem and then to define it —here negatively. The first four terms serve as partial definitions and, in fact, supplement each other. The adroit humanist can see that paralysis and death have much in common, as do knowledge and timelessness. These four, then, set the limits for the state the poem is describing and for the speaker who is describing it. The fifth option stands out as dealing with a different dimension of the poem—that of mood—and in that respect is a bad question. But you could not make a fifth and incorrect option that would not either stand out prominently or retain some defensible points; "action and inaction," "past and future," "myth and reality," "chaos and order" are all defensible; "death and rebirth," "poverty and richness," "sea and land," "stone and wood" are all ridiculous. In short, once the premise that the poem has a certain kind of resonance (from activity to inactivity and from ignorance to knowledge) is granted, any number of further resonances are equally viable. The question, then, is a good teaching question, but not a good question for objective assessment.

A second untenable question asks: Which of the following words gives the first suggestion that something unusual is going to happen? (A) "had" (line 1); (B) "cave" (line 1); (C) "sheer" (line 2); (D) "hung" (line 3); (E) "wheeled" (line 4). This question presents a similar problem, for a poem is by definition an organic unit, and an argument can be made that every word has a potency equal to any other. "Cave" does stand out as a modifier of the image, but "had" by virtue of its tense also seems to establish an ominous condition. The sensitive reader might agree that one word is more suggestive than another, but the difference is simply of degree.

Both of these questions indicate some of the limits involved in multiple-choice testing of the understanding and appreciation of literature. The literature committee and its consultants are constantly concerned with these limits and are always in search of ways by which an objective test can probe more deeply into the student's response. Other experiments have been carried out using questions that ask for best and second-best answers, that ask a student to judge the validity of certain responses, and that try to assess his ability to judge the appropriateness of various suggestions for omitted lines from a poem.

All of these experiments seek to make the test of literary acumen one which finds out how deeply students examine a poem, by following a pattern of questions similar to those asked about "Medusa." This pattern attempts to move from a test of sharpness of observation to a test of depth of understanding, the two qualities which seem to mark the mature reader of literature.